

Analysis of satisfaction and topics in repeated conversation through days

Tsunehiro Arimoto¹, Hiroaki Sugiyama¹, Masahiro Mizukami¹,
Hiromi Narimatsu¹, and Ryuichiro Higashinaka^{1,2}

¹ NTT Communication Science Laboratories, NTT Corporation

² NTT Media Intelligence Laboratories, NTT Corporation

{tsunehiro.arimoto.ub,hiroaki.sugiyama.kf,masahiro.mizukami.df,
hiromi.narimatsu.eg,ryuichiro.higashinaka.tp}@hco.ntt.co.jp

Abstract

For a dialogue system to function as a daily conversation partner, it must behave naturally not only in a single conversation but also in multiple conversations with its users. Analyzing how human satisfaction and topics change in conversations when conversations accumulate is useful for developing such systems. In this study, we analyzed multiple text-chats between two strangers for four days on a controlled schedule and revealed that their satisfaction and topic distribution depend on the length of the intervals between conversations.

1 Introduction

Chat-oriented dialogue systems are currently used for various tasks (eg., recommendation, therapy, and entertainment). While most systems assume that each user has a single conversation for a few minutes, for certain tasks, some systems assume that they have longer conversations with its users.

In order to make users have longer conversations, research for improving the naturalness or consistency of multi-turn dialogues has been actively investigated (Zhang et al., 2018). However, there are not many studies that focus on multiple conversations with its users. One of the differences between a single conversation and multiple conversations is that speakers have a short or long interval between conversations.

When an interval length between conversations is short, speakers might not care the small interval and behave as if they continue a long single conversation and might strengthen their engagement to the dialogue gradually. In contrast, when an interval between conversations is long, speakers may feel a difficulty in strengthening the engagement. If a system does not consider the effect of the length of the intervals, the system may speak to the user with wrong engagement strength that makes the user disappointed.

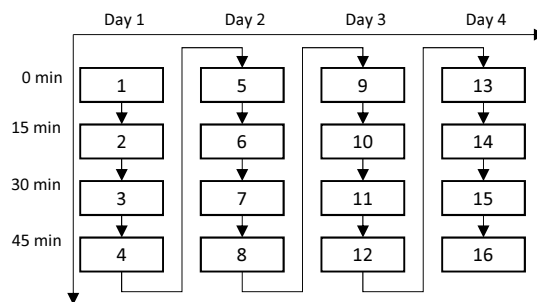


Figure 1: Scheduled collection of text chats per conversation pair (number in each cell means the order of the conversation or the cumulative number of conversations)

This study investigates how an interval length between conversations affects human satisfaction and topic selection in multiple conversations. We also analyze the trend as the number of conversation accumulates. This study focuses on dialogues in text chat in order to avoid the influence of the behavior or appearance of participants. We analyze a human-human text-chat corpus to investigate the natural behaviors of humans.

2 Repeated text-chats corpus

To investigate the effect of interval between conversations and the effect of their accumulation, we must analyze data where the time intervals between conversations are controlled. Since the level of intimacy between the speakers also affects the conversation contents, their relationships must also be controlled (Taylor and Altman, 1966).

A text-chat data collected by Higashinaka et al., to implement an interactive system satisfies these conditions (Higashinaka et al., 2014). They collected four-days long chat data from two strangers who met on text-chat.

In their data, there is a controlled time interval between each text-chat. Figure 1 shows the

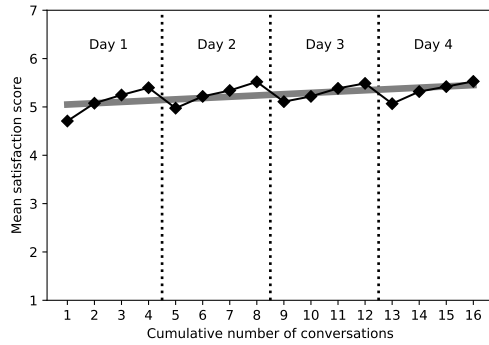


Figure 2: Satisfaction trends; Satisfaction is on a 7-Lickert scale (7 = very satisfied, 1 = completely unsatisfied). A gray line shows the regression line for all data.

recording schedule for one pair who text-chatted four times a day for four consecutive days. Because the four text-chats in each day are recorded within an hour, their corpus reveals the effect of short time intervals by comparing them. Since there is a long time interval when day changes, the comparison of two text-chats before and after the day ends reveals the effect of long time intervals. Our study analyzes part of their data: 2496 dialogues, 156 pairs, and 89 people.

3 Analysis

3.1 Human satisfaction

We analyze how human satisfaction is related to the length of interval between conversations and their accumulated amounts. We analyzed speaker satisfaction by the questionnaire results reported when they finished each text-chat.

Figure 2 shows the transition of the average scores of all the participants. Figure 2 shows that mean satisfaction score increased when the text-chats were repeated each day. Satisfaction decreased when the days changed (ex., cumulative number=4 vs number=5). These results suggest that human satisfaction increased during short intervals and decreased during long ones. The regression line for all data illustrates that human satisfaction gradually increased as the number of conversations increased.

3.2 Topic selection

Next we analyzed how topic selection changed over time. An annotator labeled the conversation topic (ex., fashion) per text-chat. The heat map in Fig. 3 shows the frequency of each topic at

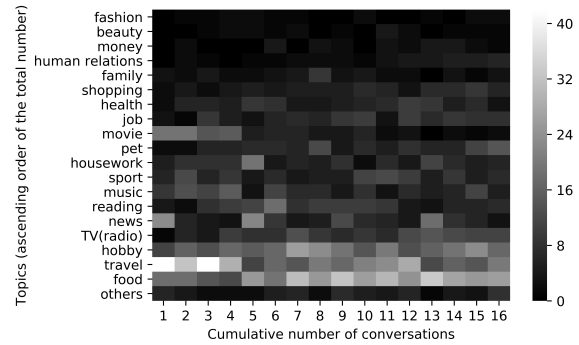


Figure 3: Number of chats on each topic with the cumulative number of conversation

each conversation timing (the cumulative number of conversation). A bright cell indicates that many conversations with its topic occur in its conversation timing. Dark cells are infrequent cells.

Distribution is uneven rather than uniform depending on each interval length and the accumulation amount of the conversations. For example, “news” often appeared after long intervals (ex., cumulative number=4 vs number=5). “Movie” often appeared on the first day (the cumulative number=1, 2, 3, 4) when the talkers are not familiar yet.

4 Conclusion

We examined the effect of multiple conversations on human satisfaction and topic selection in repeated text-chats. Our results suggest that human satisfaction and topic selection are affected by the length of the time intervals between conversations and the accumulation of the dialogue.

References

- Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. 2014. Towards an open-domain conversational system fully based on natural language processing. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 928–939.
- Dalmas A Taylor and Irwin Altman. 1966. Intimacy-scaled stimuli for use in studies of interpersonal relations. *Psychological Reports*, 19(3):729–730.
- Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*.